The Report Committee for Kristen Marie Matthews
Certifies that this is the approved version of the following report:

Reliability and Validity of the Global Physical Activity Questionnaire (GPAQ) and Its Utility: A Review of the Literature

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: _____
Xiaofen Keating

_____
Darla Castelli

Reliability and Validity of the Global Physical Activity Questionnaire (GPAQ) and Its Utility: A Review of the Literature

by

Kristen Marie Matthews, B.S.Kin.Health.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

August 2016

Dedication

For the men in my life, who have shown me what real love and courage are and for the women in my life, who have instilled in me compassion and kindness.

Acknowledgements

First, I have to thank my supervisor, Dr. Xiaofen Keating without whom this truly would not have happened. I am so grateful that you had the idea, and encouraged me, to turn an idea from a class assignment into a report. Secondly, I would like to thank my reader, Dr. Darla Castelli, for carving out the necessary time to give me such valuable feedback on this report. It would not be nearly as good without your input.

To my family and friends, I cannot thank you enough for supporting me every step of the way. This report is but one peak in a long journey, and you have been through it all with me.

To my coworkers, you have been a highlight of my time at UT. Thank you for making work such an enjoyable place to be and for always being flexible so that I could accomplish what I wanted in-, and outside, of work.

To the university, thank you for being an institution that I could change majors and career paths, but remain in a top tier program. The lessons I have learned from professors, friends, and coworkers are invaluable and made me more socially aware than I was five years ago.

Abstract

Reliability and Validity of the Global Physical Activity Questionnaire (GPAQ) and Its Utility: A Review of the Literature

Kristen Marie Matthews, M.A.

The University of Texas at Austin, 2016

Supervisor: Xiaofen Keating

Objective: To analyze the literature on the Global Physical Activity Questionnaire in regards to its reliability, validity, and utility.

Methods: A review of the literature was performed in June 2016 using the following databases: PubMed and EBSCOhost Research Databases. Articles that analyzed the reliability and validity of the GPAQ internally, in comparison to other questionnaires, or in a country context were included. The following data were coded for each article: number of participants, mean age, validity measure(s), validity, and reliability. Frequency counts and mean values of reliability and validity were calculated.

Results: Specific populations yielded different results in terms of the reliability and validity of the GPAQ. Overall, the GPAQ has been found to have similar, if not better, reliability and validity than other questionnaires that aim to measure physical activity, such as the International Physical Activity Questionnaire (IPAQ), Madras Diabetes

Research Foundation- Physical Activity Questionnaire (MPAQ), Total Energy Expenditure Questionnaire (TEEQ), etc.

Conclusions: The reliability and validity of the GPAQ are fairly acceptable in different populations. However, the GPAQ's utility can be increased if its cultural relativity is improved throughout the world.

Table of Contents

List of Tables

Introduction

Participation in physical activity on a regular basis is well documented as a critical component of a healthy lifestyle and disease prevention (Bull, Maslin, & Armstrong, 2009). Physical inactivity has been linked to many non-communicable diseases (NCDs). These diseases account for 60% of all deaths in the world (Unwin & Alberti, 2006). NCDs such as diabetes, cancer, and cardiovascular disease can be prevented or treated by a physically active lifestyle (Igwesi-Chidobe, Godfrey, & Kengne, 2015). By 2020, NCDs are expected to account for 73% of all deaths (World Health Organization, 2013). The increase in deaths related to NCDs is reflected in the youth population of United States, as this generation is the first not expected to outlive their parents if appropriate lifestyle related interventions are not implemented in a timely manner (Olshanksy et al., 2005). While developed countries are commonly believed to be mostly affected by NCDs, studies have shown an increasing effect in other countries. Nugent (2008) predicted that the cost of NCDs to developing countries would reach $84 billion by 2015. These circumstances have led to the need for a way to surveil the physical activity level of populations.

Several different physical activity questionnaires have been developed and validated for use in developed countries. These questionnaires, although valid, are limited in scope and scale. Most of them place high importance on leisure time physical activity, which is not as commonly found in developing countries. In order to accurately survey the world's populace another questionnaire had to be developed.

The Global Physical Activity Questionnaire (GPAQ) was developed by the World Health Organization (WHO) in response to a greater interest in the role of physical activity in NCD prevention (Armstrong & Bull, 2006). The aim of the GPAQ was to enhance the IPAQ (International Physical Activity Questionnaire) in cross-cultural settings. The GPAQ uses a standardized protocol as surveillance of physical activity engagement at the population level. The questionnaire covers several components of physical activity and consists of three domains; occupational, transport-related, and leisure activities (Armstrong, Bauman, & Davies, 2000). In order to draw conclusions and advise countries in the future the GPAQ has to be reliable, valid, and adaptable across the varying populations it is intended for. This paper will examine whether the GPAQ can be considered to be reliable and valid, to what extent it is culturally relevant in specific populations (mainly Asia), and whether it may be useful to generate comparable data across nations when implementing a physical activity intervention program.

Literature Search

A search using PubMed and EBSCOhost Research Databases was conducted to identify pertinent articles through June 2016 using the following keywords: *gpaq, reliability, validity, global physical activity questionnaire, WHO, and world health organization*. Articles referenced in those found in the initial search were also examined to determine inclusion. Only peer-reviewed articles were selected for review if they fulfilled the following criteria: written in English language, found via a search engine, analyzed the reliability and/or validity of the GPAQ or used the GPAQ to analyze the validity of another questionnaire.

DATA REDUCTION AND ANALYSIS

Once included, all articles were read by the author and coded for the following: number of participants, mean age, validity measure(s), validity, and reliability. Frequency counts and mean scores of reliability and validity were collected.

Table 1: Description of the included studies.

Article	Description	Main findings
Specific populations		
Au et al., 2010	Using pedometers, examined the test retest reliability and concurrent validity of the GPAQ and IPAQ long form in Vietnam.	GPAQ underestimates total physical activity for women. Participants with unstable work patterns had poor test retest reliability, suggesting these participants struggled to report consistently on their physical activity in a typical week. Inserting "typical week" instead of "last 7 days" had a negative impact on reliability and validity. Concurrent validity was moderate for those with stable work patterns, and poor for those with unstable work patterns.
Bull et al., 2009	Assessed reliability, concurrent validity, and criterion validity using data collected from nine countries using either pedometers or accelerometers. Countries included were Bangladesh, Brazil, China, Taiwan (China), Ethiopia, India, Indonesia, Japan, Portugal, and South Africa.	Reliability coefficients were of moderate to substantial strength. Concurrent validity between GPAQ and IPAQ showed moderate to substantial strength. Criterion validity was poor to fair. Observed differences between BMI, rural/urban, and between countries.

Table 1: Description of the included studies.		
Hoos et al., 2012	Using accelerometers, examined validity of the GPAQ in regards to U.S. adult Latinas. Also examined the sensitivity to intervention change (in regards to PA) to determine whether the GPAQ could evaluate the effectiveness of PA intervention programs.	Significant cross-sectional correlations between intensity levels across GPAQ domains, sensitivity to changes in vigorous PA across time points, and intervention participants were more accurate self-reporters the more they attended intervention events.
Hu et al., 2015	A longitudinal comparative study that examined reliability and relative validity of the Chinese version of the IPAQ, GPAQ, and Total Energy Expenditure Questionnaire (TEEQ).	The intra-class correlation coefficient (ICC) and 95% confidence interval for moderate-to-vigorous PA was excellent (0.81) for the GPAQ. It was rated higher than the IPAQ, but lower than the TEEQ. The ICCs of the PA items ranged from 0.67 to 0.85 for the GPAQ which again placed it higher than the IPAQ, but lower than the TEEQ.
Sitthipornvorakul et al., 2014	Examined the correlation of PA levels assessed by pedometer with those assessed by the GPAQ in a population of "healthy" office workers.	Good reliability found with the pedometer and GPAQ. Significant, but low correlation for participants 20-29 years and no correlation for those 30+ years. 65.3% underestimated their PA level using the GPAQ while 9.3% overestimated.

Table 1: Description of the included studies.		
Soo et al., 2015	Using pedometers, examined the reliability and concurrent validity of the GPAQ by comparing it with the short form IPAQ.	Overall medians for total MET minutes per week were not statistically different between tests. Moderate-intensity recreational activities and sedentary behaviors did not show high levels of repeatability. Significant relationship between the two tests in PA level classification, but both were insignificant in relation to the pedometers. In comparison to the pedometer data, 54% of participants overestimated their PA level using the GPAQ while just 8% underestimated their PA level.
Trinh et al., 2009	Using accelerometers, examined validity and reliability of the GPAQ in Vietnam. Specifically, comparing the results from the wet and dry seasons.	Short term repeatability (Spearman=0.69) during the same season, in this study the dry season, was greater than long term (0.55). The test retest reliability between seasons yielded high Spearman correlation coefficients in sedentary (0.75), moderate PA (0.61), and moderate + vigorous PA (0.61), but very low for vigorous PA (0.37). GPAQ overestimates moderate and vigorous PA compared to the accelerometer. GPAQ has similar validity correlations as other frequently used PA surveillance questionnaires.
Extended GPAQ		
Anjana et al., 2015	To examine relative validity of the Madras Diabetes Research Foundation PAQ compared to the GPAQ. Also tested construct validity and criterion validity of the Madras PAQ using triaxial accelerometers.	Fair correlation was found between the MPAQ and GPAQ (ICC=0.40). Highest correlations in sitting (0.57), MVPA recreational (0.50), work vigorous (0.46), and transport (0.44). MPAQ's Spearman's correlation coefficients for sedentary activity, MVPA and overall PA against the accelerometer were 0.48, 0.44, and 0.46 respectively. MPAQ's one month reliability was ICC=0.82.

Table 1: Description of the included studies.		
Hagstromer et al., 2005	Examined the construct and concurrent validity of the long-form IPAQ in Swedish adults using accelerometers.	Vigorous-intensity PA and total amount of PA were significantly correlated to that of the MTI activity monitor. No significant relationship between aerobic fitness and time spent in vigorous activity. No significant relationship between body fat and any PA variables from the IPAQ.
Herrmann et al., 2013	Two studies undertaken. One study required participants to take the GPAQ 3 months apart, the IPAQ, and fitness and anthropomorphic measures. All used pedometers and some used accelerometers as well. The second study had participants take the GPAQ 10 days apart.	Against measures of physical fitness, body composition, and objective (accelerometer, pedometer) and subjective measures (IPAQ) the GPAQ showed low to moderately-high validity. Long-term reliability of reporting moderate intensity activity for recreation, work, and travel is low (<.70), but short-term is higher (>.80).
Milton et al., 2011	Developed and tested new PA measure to assess respondent's eligibility for behavior change intervention. Used the GPAQ to assess concurrent validity.	Moderate correlation of 0.53 between the GPAQ and new measure for concurrent validity.
GPAQ Characteristics		

Table 1: Description of the included studies.		
Chu et al., 2015	To examine psychometric properties of self-administered with the original interviewer-administered approach. Whether different definitions of accelerometry-based PA bouts affect validity.	Reliability of total MVPA higher for self-administered group (ICC= 0.79 vs. 0.28). Fair to moderate criterion validity with accelerometry data. Overestimation of self-reported MVPA. Stronger criterion validity correlation in vigorous activities. Comparable reliability between self- and interviewer-administered.
Cleland et al., 2014	Assessed validity of the GPAQ using accelerometer data in measuring and assessing change in MVPA and sedentary behavior.	Moderate level of agreement for criterion validity for MVPA and change in MVPA. Statistically significant difference in measuring median MVPA minutes in females. Statistically significant difference in median minutes of sedentary behavior for the overall group, and male/female groups. GPAQ does not estimate an individual's level of PA well, but does well when estimating that of a group. GPAQ can appropriately assess effectiveness of PA interventions in a community or population setting.
Misra et al., 2014	Assessed repeatability of the GPAQ and validate it in comparison to the IPAQ using pedometers in India.	ICCs for GPAQ varied from 0.37-0.81, higher than the IPAQ. A retest of the GPAQ a month later yielded a Spearman's rho of 0.40-0.59 for different activities. The only portion not in this range was moderate intensity which yielded a Spearman's rho of only 0.21. The questionnaires showed a high level of concurrent validity (Spearman's Rho 0.89-1.00 and ICC 0.76-0.91). Significant agreement between self-reported PA and pedometer count, greater in females than males.

Synthesis

A total of 14 articles met the inclusion criteria and were selected for review. A narrative and descriptive form will be used to synthesize the findings of this study (see Table 1). The following discussion will focus on characteristics, validity and reliability found in different contexts, utility and modifications of the GPAQ, and implications for future research.

GPAQ Characteristics

Scope of GPAQ

The scope of the GPAQ is wider than other commonly used PA questionnaires and includes three domains instead of only leisure physical activities. Its utility is much greater than other PAQs because of its scope, measuring a wider age range and occupation.

Measurement uniqueness

The validity and reliability of a self-reported questionnaire hinge on sound measurement methods. The uniqueness of the GPAQ is worth noting because it helps readers better understand the reliability and validity of the questionnaire.

Measuring time duration

The GPAQ uses "a typical week" in its questions about PA. Instead of measuring last 7 days' physical activity, the use of "a typical week" yields data representing participants' average PA patterns. Also, the data are more reliable when PA is measured during a typical week.

The use of activity check list

The use of a pre-set activity list has pros and cons. The use of an activity list can help remind participants of activities they performed to increase reliability and validity. Moreover, it can also standardize the data collected. Because the GPAQ is designed for global use, however, the activity check list has its inherent limitation, especially when PA varies greatly in countries and/or regions. When the activity list is not inclusive, some popular activities in certain countries may be missing from the check list, resulting in unacceptable reliability and validity in some countries.

Length of GPAQ-utility and feasibility

The original version of the GPAQ consisted of 19 questions. After obtaining feedback from researchers and practitioners, a shorter second version was published. This version eliminated three questions that were deemed redundant and now is comprised of 16 questions (Armstrong & Bull, 2006). On average, on 5-10 minutes is needed to complete the questionnaire.

Administering methods

There are many methods to conduct surveys which have pros and cons. The GPAQ can be administered by others or self-administered. Interviewer administered requires a trained interviewer while self-administered may be more cost effective, if valid and reliable. It is therefore important to study possible differences in reliability and validity between the two administrations of the GPAQ. Chu et al. (2015) tested just that and found a high level of comparability between the two administrations. This is important because the self-administered test is inexpensive and this study shows that the GPAQ is robust, regardless of administering methods.

Analysis of characteristics

Four studies were found that analyze specific characteristics of the GPAQ (see Table 2). Although Cleland et al. (2014) and Misra, Upadhyay, Krishnan, Sharma, & Kapoor (2014) reported a low level of validity for individual level of MVPA, all four reported a moderate to high level of validity for a community setting MVPA. Individual levels of MVPA were typically over reported while sedentary behavior was found to be under reported in two studies. Even though Cleland et al. (2014) found lower levels of validity on the individual level, a global questionnaire such as the GPAQ is more concerned with evaluating larger populations in which setting it has an acceptable level of validity.

Studies from countries with a higher GDP reported a higher level of validity, possibly due to the higher education level of its populace (Cleland et al., 2014). Chu, Ng,

Koh, & Müller-Riemenschneider (2015) found that the GPAQ had the strongest reliability and validity when assessing vigorous recreational activities. Three studies that tested reproducibility reported the GPAQ had an acceptable level of repeatability both in the short-term and long-term testing.

VARIATIONS OF RELIABILITY AND VALIDITY IN SPECIFIC POPULATIONS

All 13 studies analyzed the validity and reliability of the GPAQ in different populations around the globe (see Table 2). It seems that the GPAQ is an appropriate questionnaire that can apply to the whole world with acceptable reliability and validity.

Methods used to test GPAQ validity

Pedometers and accelerometers were common ways to confirm criterion validity. Interestingly, accelerometers in general generate more accurate PA data than pedometers do (Soo et al., 2015). Due to its high costs, the sample sizes using accelerometers were usually much smaller than those using pedometers (Misra et al., 2014; Sitthipornvorakul et al., 2014). However, the validity results using both pedometers and accelerometers were about the same (see Table 2).

Variations of validity

One large obstacle the GPAQ faces is surveilling countries that have a wet and dry season, which generates a great impact on PA levels (Trinh, Nguyen, van der Ploeg, Dibley, & Bauman, 2009). PA patterns, especially in the occupation domain, can vary drastically from season to season, suggesting only certain types of occupation activities

would occur because of the weather. For example, in Vietnam occupational physical activity can vary drastically in rural areas as seasons change and crops grow. All of the studies that analyzed concurrent validity found an acceptable consistency. However, one study found that when assessing PA in countries with a dry and wet season concurrent validity was lower during the wet season (Au et al., 2010). The GPAQ does an acceptable job accounting for these differences, but could be improved when compared to other questionnaires that are narrower in scope.

Several studies found over-reporting of vigorous PA and under-reporting of the lower levels of PA. The over- and under-reporting of PA causes the GPAQ to overestimate the number of active adults; however that overestimation was deemed typical (Soo, Wan Abdul Manan, & Wan Suriati, 2015). There was a high level of correlation between the GPAQ and pedometers when observing total amount of PA. Hoos, Marshall, & Arredondo (2012) reported that the GPAQ had a significant correlation of sensitivity to change in PA.

One of the primary findings of this study was that when the GPAQ is analyzed on a country level each country has a different result. These differences are evident in Cleland et al. (2014) and Herrmann et al. (2013) which found substantially higher correlations for minutes of MVPA in the U.S.A. (r=0.26) and the UK (r=0.484) than found by Bull et al. (2009) in lower income countries South Africa (r=-0.03) and China (r=0.23). This difference may be due to wealth and education. It may be necessary to have further explanation of terms for participants in lower income areas (i.e. what is a moderate-intensity activity versus a vigorous one?).

The GPAQ has been analyzed as a possible form to assess effectiveness of PA intervention programs. The results have been mixed with Hoos, Marshall, & Arredondo (2012) finding that the GPAQ was sensitive to changes in vigorous PA across time points, however Trinh et al. (2009) found it was not a good measure for assessing the effectiveness of intervention programs. Trinh et al. (2009) argued that because the GPAQ was not sensitive to change in seasons (wet/dry) it was not suitable for intervention programs. Further research is needed to better determine whether the GPAQ is a good measure of effectiveness of PA intervention programs.

Reliability

When examining the internal consistency of the GPAQ several studies noted the influence of the change of seasons much like they did for validity. Due to very extreme seasons the GPAQ was found to be more reliable when the retest happened in the short-term (Trinh et al., 2009). With long-term testing the GPAQ did not do as well largely in part to possible seasonal changes. If the weather changes are truly to blame for this decrease in repeatability it does bolster the repeatability of the GPAQ in countries that do not experience vast changes due to monsoon season or other phenomena.

Bull, Maslin, & Armstrong (2009) reported stronger reliability coefficients in males versus females and urban settings versus rural settings. More study would need to be conducted into these differences. Similar to the studies' findings regarding validity they found that repeatability was strong for items relating to total PA, but low for moderate to intense PA as well as sedentary behavior (Hoos, Espinoza, Marshall, &

Arredondo, 2012). One possible cause for this is the lack of understanding what constitutes as moderate to vigorous. There may need to be further clarification in the GPAQ, possibly using examples, in order to increase participants' understanding when answering questions related to this topic.

Concurrent Validity

There have been many PAQs developed before and after the GPAQ was developed. One such PAQ is the MPAQ (Madras Diabetes Research Foundation-Physical Activity Questionnaire). It is designed solely for epidemiological studies in India. Anjana et al. (2015) found the correlation between the GPAQ and MPAQ to be fair (r=0.40 overall). One of the MPAQ's strengths is its ability to account for the variability of PA during the dry and wet season by capturing details of up to two jobs and elicits time spent sitting, standing, walking, and climbing in each job. In its current state, the GPAQ is limited to capturing information in one job and does not distinguish between sitting, standing, walking, and climbing. The MPAQ is also more culturally relevant to its target populace which could be a reason there is not a higher level of correlation between the two. For example, the MPAQ specifies about the seasonality and nature of occupations performed. The GPAQ has a general question in the transport domain that groups walking and cycling together while the MPAQ is more culturally relevant by splitting the two. These are two important distinctions between questionnaires that are global versus local. If possible the GPAQ needs to attempt to make allowances for more cultural inclusivity and account for changes in seasons.

One of the most prominent PAQ that is on a similar scale as the GPAQ is the IPAQ. A group of experts in the field collaborated to develop the IPAQ in 1998 in order to validly and reliably measure PA across many populations (Hagstromer, Oja, & Sjostrom, 2005). Herrmann, Heumann, Der Ananian, & Ainsworth (2013) found that the reliability and validity of the GPAQ is comparable to that of the IPAQ. However, the GPAQ has the added bonus of collecting information in three specific domains (occupational, transport, and leisure). Hu et al. (2015) conducted a study in China that found the GPAQ had higher reliability and validity. It is possible that this difference is due to cultural differences which indicate more research should be conducted to better understand this relationship.

Table 2: Summary of review findings.

References	Location	Participants	Mean Age(SD)	Validity measure(s)	Validity	Reliability	Notes
Anjana et al., 2015	India	103	44(14) & 42(13)	Accelerometers and MPAQ	Overall fair correlation (ICC=0.40) Highest for sitting (0.57), MVPA recreational (0.50), work vigorous (0.46), and transport (0.44)	N/A	Validity study (mean age urban and rural, respectively)
Au et al., 2010	Vietnam	251	25-64	IPAQ	Stable work patterns: moderate Unstable work patterns: poor	Stable: r=0.39 Unstable: r=-0.05	Pedometers were used, but not as criterion measure
Bull et al., 2009	Bangladesh, Brazil, China, Taiwan (China), Ethiopia, India, Indonesia, Japan, Portugal, and South Africa	2657	18-75	Pedometers, accelerometers, and IPAQ	Concurrent: moderate to strong (0.45-0.65) Criterion: poor-fair (0.06-0.35)	Moderate to substantial (Kappa 0.67-0.73, r=0.67-0.81)	1st iteration of the GPAQ

Table 2: Summary of review findings.

Chu et al., 2015	Singapore	108	31(26.8-47.3)	Accelerometers	Fair to moderate for MVPA in self (r=0.3) and interviewer (r=0.46)	MVPA was moderate for self (r=0.61) vs. interviewer (r=0.63)	
Cleland et al., 2014	UK	95	44(14)	Accelerometers	Criterion: Minutes of MVPA (r=0.484) Minutes of sedentary behavior (r=0.187)	N/A	Validation study
	UK	30	46(13)	N/A	N/A	Poor agreement over time points (r=-0.024)	Test-retest study
Herrmann et al., 2013	U.S.A.	69	43.1(11.4)	Pedometers, accelerometers, and IPAQ	Overestimates at lower levels of activity and underestimates as more activity is reported	Three month test-retest (r=0.53-0.83)	Validity and long-term test-retest reliability
	U.S.A.	16	40.2(12.6)	Pedometers and accelerometers	N/A	10 day test-retest (r=0.83-0.96)	Short-term test-retest reliability

Table 2: Summary of review findings.

Hoos et al., 2012	U.S.A.	72	43.01(9.05)	Accelerometers and self-reported vigorous PA	Significant correlation of sensitivity to change (r=0.383)	N/A	Latinas (female)
Hu et al., 2015	China	205	51.36(10.25)	PA-log, IPAQ-S-C, and TEEQ-C	Higher reliability and validity than IPAQ, lower than TEEQ ICCs IPAQ: 0.51-0.8 GPAQ: 0.67-0.85 TEEQ: 0.74-0.94		
Milton et al., 2011	England, Scotland, and Wales	480	18-64	GPAQ and single-item measure developed for study	Modest (r=0.53)	N/A	Used GPAQ to assess APS validity
Misra et al., 2014	India	234	15-65	Pedometers and IPAQ	Reasonable agreement r= >0.9 ICC: 0.76-0.91	r=0.4-0.59 ICC 0.56-0.68	
Sitthipornvorakul et al., 2014	Thailand	320	34.8(6.2)	Pedometers	Criterion: Physical activity level (r=0.08)	Kendall's tau-b=0.89	Office workers

Table 2: Summary of review findings.

Soo et al., 2015	Malaysia	100	36.4(10.5)	Pedometers and IPAQ-S	Concurrent: moderate associations for total, vigorous, moderate, and sedentary (r=0.309-0.466) Criterion: low, but significant between total PA (GPAQ) and average steps/day (pedometer) (r=0.265)	Two week test-retest Total MET/wk and 4 domains were not significantly different Moderate recreational activities (z=-3.515) and sedentary (z=-3.272) were poor
Trinh et al., 2009	Vietnam	169	44.7(11.1)	Accelerometers	Criterion: Dry season (r=0.34) Wet season (r=0.20)	Two week test-retest r=0.69 Two month test-retest r=0.55

Discussion

The importance of monitoring PA levels on a regular basis in the general population cannot be overstated because it has the potential to save billions of dollars in medical treatments caused by sedentary lifestyles and increase quality of life. Measurements with sound reliability and validity are the first step to ensure effective assessment of PA. Although objective PA devices have become available in recent years, improving the accuracy of PA data, self-reported questionnaire/surveys still have its superior advantages due to its low cost and capability of reaching much larger samples of populations, generating the best cost-effect ratio. More research on the utility of self-reported questionnaires is needed in the future.

Overall, there are multiple ways to confirm the reliability and validity of self-reported physical activity. The purpose of this study was to analyze the accuracy of the GPAQ to measure an individual's PA regardless of age and location. Through the examination of previously published articles on the topic, it was discovered that pedometers and accelerometers were common ways to confirm criterion validity. The current project highlights three issues related to the revalidations of the GPAQ: (a) the use of relatively objective PA measuring device for validations in free living settings; (b) the impact of cultural differences; and (c) the role of education level.

USE OF PEDOMETERS AND ACCELEROMETERS

Six studies employed the use of pedometers for all or most of their participants. The data collected by the pedometers was used as an objective criterion measure of

physical activity to which the GPAQ estimates could be compared. Although studies have shown that pedometers provide a valid and accurate measure of PA and are considered useful in PA studies, there are significant limitations to these devices. Pedometers are unable to discern the intensity of an activity which would be helpful when analyzing intensity specific estimates the GPAQ provides. Studies have found that pedometers may not accurately record steps for people with abnormal gait patterns and people that are obese. Pedometers also may not accurately capture activities in which the lower body is stationary (i.e. pushing, lifting, and carrying). These factors could contribute to studies that have found the GPAQ overestimates PA compared to the pedometers, because the pedometers may be underestimating PA (Soo et al., 2015). Eight studies used accelerometers. While accelerometers are considered to be more accurate than pedometers, they may still underestimate upper body movement and movements with a weak vertical component like cycling (Trinh et al. 2009). At this time there is no "gold standard" technology to assess free-living PA.

GLOBAL CULTURAL RELEVANCE

Any global questionnaire will have its limitations. One of the largest for the GPAQ is being standard across the globe, yet maintaining relevance to the participants. A questionnaire that is developed for a specific country can be more culturally specific than a global questionnaire can. However, the GPAQ has been compared to some of these one country questionnaires and has shown some correlation. The GPAQ must continue to limit the culture gap as much as possible in order to maintain relevance as a global

questionnaire. Some questions may need to be broken up into more specific questions to increase the validity and reliability. Part of this cultural relevance is accounting for seasonal changes in physical activity. There was much discussion in this paper and the studies associated with it that the GPAQ struggled with acknowledging changes in occupational physical activity during the wet and dry season in Southeast Asia. Adaptations that would account for those changes would greatly increase the short-term and long-term repeatability in these populations and others similar to it.

IMPLICATIONS FOR FUTURE RESEARCH

The original intent of the GPAQ was to survey people around the globe to discover their physical activity behaviors and attempt to observe a link between that and their rates of NCDs. Two studies (Hoos et al. & Trinh et al.) attempted to use the GPAQ to determine the effectiveness of a PA based intervention program. Although this was not the original intended use, if it can be used to determine effectiveness it could be very useful in the application of PA based intervention programs. Hoos et al. and Trinh et al. came to different conclusions on their work in this topic so further research needs to be conducted to determine whether the GPAQ can be utilized in this manner.

References

Anjana, R.M., Sudha, V., Lakshmipriya, N., Subhashini, S., Pradeepa, R., Geetha, L., . . . Mohan, V. (2015). Reliability and validity of a new physical activity questionnaire for India. *International Journal of Behavioral Nutrition & Physical Activity,* 12(1), 1-12.

Armstrong, T., Bauman, A., & Davies, J. (2000). Physical activity patterns of Australian adults. Cat. No. CVD 10. Canberra: AIHW.

Armstrong, T., & Bull, F. (2006). Development of the world health organization Global Physical Activity Questionnaire (GPAQ). *Journal of Public Health,* 14(2), 66-70.

Au, T.B., Blizzard, L., Schmidt, M., Luc, P.H., Magnussen, C., & Dwyer, T. (2010). Reliability and Validity of the Global Physical Activity Questionnaire in Vietnam. *Journal of Physical Activity & Health,* 7(3), 410-18.

Bull, F.C., Maslin, T.S., & Armstrong, T. (2009). Global Physical Activity Questionnaire (GPAQ): Nine Country Reliability and Validity Study. *Journal of Physical Activity & Health,* 6(6), 790-804.

Chu, A.H.Y., Ng, S.H.X., Koh, D., & Müller-Riemenschneider, F. (2015). Reliability and validity of the self- and interviewer-administered versions of the Global Physical Activity Questionnaire (GPAQ). *PLoS ONE,* 10(9).

Cleland, C.L., Hunter, R.F., Kee, F., Cupples, M.E., Sallis, J.F., & Tully, M.A. (2014). Validity of the Global Physical Activity Questionnaire (GPAQ) in assessing levels and change in moderate-vigorous physical activity and sedentary behavior. *BMC Public Health,* 14.

Hagstromer, M., Oja, P., & Sjostrom, M. (2005). The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. *Public Health Nutrition,* 9(6), 755-62.

Herrmann, S.D., Heumann, K.J., Der Ananian, C.A., & Ainsworth, B.E. (2013). Validity and Reliability of the Global Physical Activity Questionnaire (GPAQ). *Measurement in Physical Education & Exercise Science,* 17(3), 221-35.

Hoos, T., Espinoza, N., Marshall, S., & Arredondo, E.M. (2012). Validity of the Global Physical Activity Questionnaire (GPAQ) in adult Latinas. *Journal of Physical Activity & Health,* 9(5), 698-705.

Hu, B., Lin, L.F., Zhuang, M.Q., Yuan, Z.Y., Li, S.Y., Yang, Y.J., . . . Lu, M. (2015). Reliability and relative validity of three physical activity questionnaires in Taizhou population of China: the Taizhou Longitudinal Study. *Public Health,* 129(9), 1211-7.

Igwesi-Chidobe, C., Godfrey, E., & Kengne, A. (2015). Effective components of exercise and physical activity-related behavior-change interventions for chronic non-communicable diseases in Africa: protocol for a systematic mixed studies review with meta-analysis. *British Medical Journal Open,* 5(8).

Milton, K., Bull, F.C., & Bauman, A. (2011). Reliability and validity testing of a single-item physical activity measure. *British Journal of Sports Medicine,* 45(3), 203-208.

Misra, P., Upadhyay, R.P., Krishnan, A., Sharma, N., & Kapoor, S.K. (2014). A Community Based Study to Test the Reliability and Validity of Physical Activity Measurement Techniques. *International Journal of Preventive Medicine,* 5(8), 952-59.

Nugent, R. (2008). Chronic diseases in developing countries: health and economic burdens. *Annals of the New York Academy of Sciences,* 1136, 70-79.

Olshansky, S.J., Passaro, D.J., Hershow, R.C., Layden, J., Carnes, B.A., Brody, J., . . . Ludwig, D.S. (2005). A potential decline in life expectancy in the United States in the 21st century. *New England Journal of Medicine,* 352(11), 1138-45.

Sitthipornvorakul, E., Janwantanakul, P., & van der Beek, A.J. (2014). Correlation between pedometer and the Global Physical Activity Questionnaire on physical activity measurement in office workers. *BMC Research Notes*.

Soo, K.I., Wan Abdul Manan, Wm., & Wan Suriati, Wn. (2015). The Bahasa Melayu Version of the Global Physical Activity Questionnaire: Reliability and Validity Study in Malaysia. *Asia-Pacific Journal of Public Health,* 27(2), 184-93.

Trinh, O.T.H., Nguyen, N.D., van der Ploeg, H.P., Dibley, M.J., & Bauman, A. (2009). Test-Retest Repeatability and Relative Validity of the Global Physical Activity Questionnaire in a Developing Country Context. *Journal of Physical Activity & Health,* 6, S46-53.

Unwin, N., & Alberti, K.G. (2006). Chronic non-communicable diseases. *Annals of Tropical Medicine and Parasitology,* 100(5-6), 455-64.

World Health Organization: Global action plan for the prevention and control of noncommunicable diseases 2013-2020. In. Edited by WHO. Geneva, Switzerland: WHO; 2013: 55.